

CLAIMS

1. Method for obtaining protein chains constituting the extracellular
5 haemoglobin molecule of Annelida, in particular of *Arenicola marina*,

said method being characterized in that it comprises a stage of bringing together a
sample of extracellular haemoglobin of Annelida, in particular of *Arenicola marina* and
at least one dissociating agent, and if appropriate a reducing agent, in particular a
mixture made up of dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine hydrochloride
10 (TCEP) or beta-mercaptoethanol and a dissociation buffer, for a sufficient time to
separate the protein chains from each other.

2. Method according to claim 1, characterized in that the dissociation buffer
comprises a buffering agent at a concentration comprised between approximately 0.05
15 M and approximately 0.1 M, in particular Trisma (tris[hydroxymethyl]aminomethane),
and 0 to 10 mM of EDTA adjusted to a pH comprised between approximately 5 and
approximately 12, and preferably between approximately 7.5 and 12

3. Method according to claim 1 or 2, characterized in that the protein chains
20 constituting said molecule are obtained by the reduction of four sub-units by a reducing
agent, for example in the presence of DTT, said sub-units themselves being obtained by
bringing together a sample of extracellular haemoglobin of *Arenicola marina* and
different dissociating agents, in particular a dissociation buffer.

4. Method for preparing primer pairs from the protein chains as obtained
25 according to the method as defined in one of claims 1 to 3, said method being
characterized in that it comprises the following stages:

- the isolation of each of the protein chains constituting the haemoglobin
molecule as obtained according to the method according to one of claims 1 to 3,
- 30 – the microsequencing of each of the abovementioned isolated protein chains by
mass spectrometry and Edman sequencing, in order to obtain a microsequence
corresponding to each of the sequences made up of 5 to 20 amino acids, and
- the determination of degenerated primer pairs from the abovementioned
microsequences.

5. Primer pairs as obtained according to the method of claim 4, said pairs being in particular the following:

- 5 a) *Sense primer:* GAR TGY GGN CCN TTR CAR CG (SEQ ID NO: 21)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
- b) *Sense primer:* TGY GGN ATH CTN CAR CG (SEQ ID NO: 23)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
- 10 c) *Sense primer:* AAR GTI AAR CAN AAC TGG (SEQ ID NO: 24)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
- d) *Sense primer:* TGY TGY AGY ATH GAR GAY CG (SEQ ID NO: 25)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
- e) *Sense primer:* AAR GTN ATH TTY GGN AGR GA (SEQ ID NO: 26)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
- 15 f) *Sense primer:* GAR CAY CAR TGY GGN GGN GA (SEQ ID NO: 27)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)

where:

R represents A or G,

Y represents C or T,

20 N represents A, G, C or T,

I represents inosine,

H represents A, C or T

25 6. Method for preparing nucleotide sequences encoding the protein chains constituting the haemoglobin molecule of *Arenicola marina*, from the primers as obtained according to the method of claim 4, said method being characterized in that it corresponds to a polymerase chain amplification method (PCR), comprising a repetition of at least 30 times the cycle constituted by the following stages:

- 30 * the denaturation, by heating, of the monocatenary cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, so as to denature any secondary structures and RNA residuals, said cDNA being obtained from mRNA, this stage making it possible to obtain strands of denatured monocatenary cDNA,

* the hybridization of the primer pairs as obtained by the method as defined above to the strands of abovementioned denatured monocatenary cDNA at an appropriate temperature, in order to obtain hybridized primers, and

5 * the synthesis of the complementary strand of the cDNA by a polymerase at an appropriate temperature, from the hybridized primers as obtained in the preceding stage.

7. Preparation method according to claim 6, characterized in that:

10 – the first stage of said method is a stage of denaturation of approximately 10 seconds to approximately 5 minutes at a temperature comprised between approximately 90°C and approximately 110°C,

 – the cycle, repeated approximately 30 to 40 times, comprises the following stages:

15 * a stage of denaturation of approximately 10 seconds to approximately 5 minutes, at a temperature comprised between approximately 90°C and approximately 110°C,

 * a stage of hybridization of approximately 20 seconds to approximately 2 minutes, at a temperature comprised between approximately 50°C and approximately 60°C, and preferably between approximately 50°C and approximately 56°C,

20 * a stage of elongation of approximately 20 seconds to approximately 1 minute and 30 seconds, at a temperature comprised between approximately 70°C and approximately 75°C, and

25 – the last stage of the method is a stage of elongation of approximately 5 minutes to approximately 15 minutes at a temperature comprised between approximately 70°C and approximately 75°C.

8. Preparation method according to claim 6 or 7, characterized in that the primer pairs used are as defined in claim 5.

30

9. Preparation method according to claim 8, characterized in that the primer pair used is the following: (GAR TGY GGN CCN TTR CAR CG; CTC CTC TCC TCT CCT CTT CCT), R, Y and N being as defined in claim 5,

said method being characterized in that:

- the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,
- the cycle, repeated 35 times, comprises the following stages:
 - * a stage of denaturation of 30 seconds at a temperature equal to 95°C,
 - * a stage of hybridization of 30 seconds at a temperature equal to 56°C,
 - * a stage of elongation of 40 seconds at a temperature equal to 72°C, and
- the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 13,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 1.

10. Preparation method according to claim 8, characterized in that the primer pair used is the following: (TGY GGN ATH CTN CAR CG; CTC CTC TCC TCT CCT CTT CCT), N, Y and R being as defined in claim 5, said method being characterized in that:

- the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,
- the cycle, repeated 35 times, comprises the following stages:
 - * a stage of denaturation of 30 seconds at a temperature equal to 95°C,
 - * a stage of hybridization of 30 seconds at a temperature equal to 53°C,
 - * a stage of elongation of 40 seconds at a temperature equal to 72°C, and
- the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 15,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 3.

11. Preparation method according to claim 8, characterized in that the primer pair used is the following: (AAR GTI AAR CAN AAC TGG; CTC CTC TCC TCT CCT CTT CCT), R, I and N being as defined in claim 5, said method being characterized in that:

– the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

- * a stage of denaturation of 1 minute at a temperature equal to 95°C,
- * a stage of hybridization of 1 minute at a temperature equal to 50°C,
- * a stage of elongation of 1 minute and 30 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 17,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 5.

12. Preparation method according to claim 8, characterized in that the primer pair used is the following: (TGY TGY AGY ATH GAR GAY CG; CTC CTC TCC TCT CCT CTT CCT), Y, H and R being as defined in claim 5, said method being characterized in that:

– the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

- * a stage of denaturation of 30 seconds at a temperature equal to 95°C,
- * a stage of hybridization of 40 seconds at a temperature equal to 52°C,
- * a stage of elongation of 30 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 19,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 7.

13. Preparation method according to claim 8, characterized in that the primer pair used is the following: (AAR GTN ATH TTY GGN AGR GA; CTC CTC TCC TCT CCT CTT CCT), R, H, N and Y being as defined in claim 5, said method being characterized in that:

– the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

* a stage of denaturation of 30 seconds at a temperature equal to 95°C,

5 * a stage of hybridization of 40 seconds at a temperature equal to 52°C,

* a stage of elongation of 30 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain a partial reference nucleotide sequence,

10 said method comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 9.

14. Preparation method according to claim 8, characterized in that the primer pair used is the following: (GAR CAY CAR TGY GGN GGN GA, CTC CTC TCC TCT CCT CTT CCT), R, N and Y being as defined in claim 5,

15

said method being characterized in that:

– the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

20 * a stage of denaturation of 40 seconds at a temperature equal to 95°C,

* a stage of hybridization of 1 minute at a temperature equal to 58°C,

* a stage of elongation of 1 minute and 10 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

25

in order to obtain a partial reference nucleotide sequence,

said method comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 11.

30 15. Proteins encoded by one of the sequences as obtained according to the method as defined in any one of claims 6 to 14.

16. Protein according to claim 15, characterized in that it comprises or is constituted by:

- the sequence SEQ ID NO: 2,
- or any sequence derived from the sequence SEQ ID NO: 2 or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,
- or any sequence homologous to the sequence SEQ ID NO: 2 or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 2, provided that said homologous sequence allows the transport of oxygen,
- or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 2.

17. Protein according to claim 15, characterized in that it comprises or is constituted by:

- the sequence SEQ ID NO: 4,
- or any sequence derived from the sequence SEQ ID NO: 4, or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,
- or any sequence homologous to the sequence SEQ ID NO: 4, or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 4, provided that said homologous sequence allows the transport of oxygen,
- or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 4.

18 Protein according to claim 15, characterized in that it comprises or is constituted by:

- the sequence SEQ ID NO: 6,

– or any sequence derived from the sequence SEQ ID NO: 6, or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 6, or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 6, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 6.

19. Protein according to claim 15, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 8,

– or any sequence derived from the sequence SEQ ID NO: 8, or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 8, or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 8, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 8.

20. Protein according to claim 15, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 10,

– or any sequence derived from the sequence SEQ ID NO: 10, or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 10, or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 10, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 10.

21. Protein according to claim 15, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 12,

– or any sequence derived from the sequence SEQ ID NO: 12, or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the combination of globin chains with each other,

– or any sequence homologous to the sequence SEQ ID NO: 12, or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 12, provided that said homologous sequence allows the combination of globin chains with each other,

– or any fragment of one of the sequences defined above, provided that said fragment allows the combination of globin chains with each other, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 280 contiguous amino acids in the sequence SEQ ID NO: 12.

22. Nucleotide sequences as obtained according to the method as defined in any one of claims 6 to 14.

23. Nucleotide sequences encoding a protein as defined in one of claims 15 to 21.

24 Nucleotide sequence according to claim 23, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 1 encoding SEQ ID NO: 2,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 1, and encoding a protein represented by SEQ ID NO: 2,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 1 encoding a protein derived from SEQ ID NO: 2,

– or any nucleotide sequence homologous to SEQ ID NO: 1, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 1,

– or any fragment of the nucleotide sequence SEQ ID NO: 1 or of the nucleotide sequences defined above, said fragment preferably being constituted by at least approximately 180 nucleotides, and in particular by at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

25. Nucleotide sequence according to claim 23, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 3 encoding SEQ ID NO: 4,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 3, and encoding a protein represented by SEQ ID NO: 4,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 3 encoding a protein derived from SEQ ID NO: 4,

– or any nucleotide sequence homologous to SEQ ID NO: 3, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 3,

– or any fragment of the nucleotide sequence SEQ ID NO: 3 or of the nucleotide sequences defined above, said fragment preferably being constituted by at least approximately 180 nucleotides, and in particular by at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

5 **26.** Nucleotide sequence according to claim 23, characterized in that it comprises or is constituted by:

- the nucleotide sequence SEQ ID NO: 5 encoding SEQ ID NO: 6,
- or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 5, and encoding a protein represented by SEQ ID NO: 6,
- 10 – or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 5 encoding a protein derived from SEQ ID NO: 6,
- or any nucleotide sequence homologous to SEQ ID NO: 5, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 5,
- 15 – or any fragment of the nucleotide sequence SEQ ID NO: 5 or of the nucleotide sequences defined above, said fragment preferably being constituted by at least approximately 180 nucleotides, and in particular by at least approximately 480 contiguous nucleotides in said sequence,
- or any nucleotide sequence complementary to the abovementioned sequences or fragments,
- 20 – or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

25 **27.** Nucleotide sequence according to claim 23, characterized in that it comprises or is constituted by:

- the nucleotide sequence SEQ ID NO: 7 encoding SEQ ID NO: 8,
- or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 7, and encoding a protein represented by SEQ ID NO: 8,
- 30 – or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 7 encoding a protein derived from SEQ ID NO: 8,

– or any nucleotide sequence homologous to SEQ ID NO: 7, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 7,

– or any fragment of the nucleotide sequence SEQ ID NO: 7 or of the nucleotide sequences defined above, said fragment preferably being constituted by at least approximately 180 nucleotides, and in particular by at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

28. Nucleotide sequence according to claim 23, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 9 encoding SEQ ID NO: 10,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 9, and encoding a protein represented by SEQ ID NO: 10,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 9 encoding a protein derived from SEQ ID NO: 10,

– or any nucleotide sequence homologous to SEQ ID NO: 9, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 9,

– or any fragment of the nucleotide sequence SEQ ID NO: 9 or of the nucleotide sequences defined above, said fragment preferably being constituted by at least approximately 180 nucleotides, and in particular by at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

29. Nucleotide sequence according to claim 23, characterized in that it comprises or is constituted by:

- the nucleotide sequence SEQ ID NO: 11 encoding SEQ ID NO: 12,
- 5 – or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 11, and encoding a protein represented by SEQ ID NO: 12,
- or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 11 encoding a protein derived from SEQ ID NO: 12,
- 10 – or any nucleotide sequence homologous to SEQ ID NO: 11, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 11,
- or any fragment of the nucleotide sequence SEQ ID NO: 11 or of the nucleotide sequences defined above, said fragment preferably being constituted by at least approximately 180 nucleotides, and in particular by at least approximately 800
- 15 contiguous nucleotides in said sequence,
- or any nucleotide sequence complementary to the abovementioned sequences or fragments,
- or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or
- 20 fragments.

30. Preparation method according to claim 6, for nucleotide sequences encoding the protein chains constituting the haemoglobin molecule of Annelida, in particular of *Arenicola marina*, said method being characterized in that it comprises the following

25 stages:

- a stage of bringing together the abovementioned haemoglobin molecule with at least one dissociating agent and a reducing agent, in particular a mixture made up of dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine hydrochloride (TCEP) or beta-mercaptoethanol and a dissociation buffer, for a sufficient time to separate the protein
 - 30 chains from each other,
- allowing the dissociation, then the reduction of said haemoglobin molecule, in order to obtain the protein chains constituting said molecule,
- the isolation of each of the abovementioned protein chains,

– the microsequencing by mass spectrometry and Edman sequencing of each of the abovementioned isolated protein chains, in order to obtain a microsequence corresponding to each of the sequences made up of 5 to 20 amino acids,

– the determination of the degenerated primer pairs from the abovementioned microsequences,

– the preparation of the nucleotide sequences encoding the abovementioned protein chains, from the primers as obtained previously, by a polymerase chain amplification method (PCR), comprising the following stages:

– the first stage of said method is a stage of denaturation of approximately 10 seconds to approximately 5 minutes at a temperature comprised between approximately 90°C and approximately 110°C,

– the cycle, repeated approximately 30 to 40 times, comprises the following stages:

* a stage of denaturation of approximately 10 seconds to approximately 5 minutes, at a temperature comprised between approximately 90°C and approximately 110°C,

* a stage of hybridization of approximately 20 seconds to approximately 2 minutes, at a temperature comprised between approximately 50°C and approximately 56°C,

* a stage of elongation of approximately 20 seconds to approximately 1 minute and 30 seconds, at a temperature comprised between approximately 70°C and approximately 75°C, and

– the last stage of the method is a stage of elongation of approximately 5 minutes to approximately 15 minutes at a temperature comprised between approximately 70°C and approximately 75°C.